2013

Investigation of the Utility of Center Frequency in Electroencephalographic Classification of Cognitive Workload Transitions

Jones Melissa
Wright State University

Investigation of the Utility of Center Frequency in Electroencephalographic Classification of Cognitive Workload Transitions

A thesis submitted in partial fulfillment of the
requirements for the degree of
Master of Science in Engineering

By

MELISSA ANNE JONES
B.S.B.E., Wright State University, 2011

2013
Wright State University

WRIGHT STATE UNIVERSITY

GRADUATE SCHOOL

May 7, 2013

I HEREBY RECOMMEND THAT THE THESIS PREPARED UNDER MY SUPERVISION BY Melissa Anne Jones ENTITLED Investigation of the Utility of Center Frequency in Electroencephalographic Classification of Cognitive Workload Transitions BE ACCEPTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF Master of Science in Engineering.

Chandler Phillips, M.D.
Thesis Director

Thomas Hangartner, Ph.D., Chair
Department of Biomedical, Industrial and Human Factors Engineering

Committee on Final Examination

Chandler Phillips, M.D.

James Christensen, Ph.D.

David Reynolds, Ph.D.

R. William Ayres, Ph.D.
Interim Dean, Graduate School

Abstract

Jones, Melissa Anne. M.S.Egr., Department of Biomedical, Industrial, and Human Factors Engineering, Wright State University, 2013. Investigation of the Utility of Center Frequency in Electroencephalographic Classification of Cognitive Workload Transitions.

Successful classification of human cognitive workload is a vital component in identifying and avoiding potential performance deficits resulting from operator work overload. Previous research suggests that electroencephalogram (EEG) derived features, including center frequency, provide a robust signal which may be used to obtain highly accurate workload classification. The purpose of this work is to investigate evidence of physiological hysteresis and determine if center frequency improves a classifier's ability to correctly identify workload level. Results confirmed that including spectral data creates the most robust feature sets, while center frequency across all bands is equally reliable for classifying workload in the case where cognitive workload level transitions from hard to easy. There is also evidence of physiological signal asymmetry based on transition direction. In summation; spectral features are traditionally most dependable for classifying workload, however center frequency across all bands is an equally viable option for feature representation.

Table of Contents

Page

List of Figures

List of Tables

Table Page

Acknowledgements

First, I would like to thank my advisor, Dr. Chandler Phillips, for all of his support and guidance during this entire process as well as for the assistance in gaining funding for graduate school. Additionally, I would like to thank Dr. James Christensen for all his support, both academically and emotionally as I worked through the many challenges I encountered along the way. Thank you to Dr. David Reynolds for serving on my committee. Also, I would like to thank Samantha Klosterman, Justin Estepp and Jason Monnin, whom I had the pleasure to work with at Wright-Patterson Air Force Base, for continually offering advice and answering any questions I had. I am very grateful to the Consortium Research Fellows Program for the opportunity they have given me by funding my research.

Finally, I would like to thank my family and friends for their support and patience as I worked towards my goals, even though they often changed. My parents and friends were always there when I got frustrated or discouraged to cheer me on and help remind me of the end goal. They inspired me to keep working hard and continued to motivate me, even when I struggled to maintain the self – discipline to see past whatever issues I was having. Together, everyone played a big role in my success during this process, and I am appreciative of every one of you!

1. Introduction

Cognitive workload assessment performed using various physiological measures may be used to accurately determine changing levels of operator functional state in multi-task environments (Parasuraman & Wilson, 2008). Of particular interest, electroencephalography (EEG) provides continuous, highly sensitive information regarding the mental effort of an operator; however, the underlying signal may be overwhelmed by noise. Machine learning techniques are one method of combining multiple noisy inputs to reveal physiological trends in the fundamental signal (Gevins A. , et al., 1998). The term, workload transition is a widely accepted term that refers to a human operator's cognitive response to a perceived change in task difficulty. While previous work (Krulewitz, Warm, & Wohl, 1975) has explored the performance effects of human workload transitions, this work aims to expand the current understanding. Determination of center frequency for various EEG frequency bands both alone, and in conjunction with traditional spectral features, may provide additional insight into the functional state of an operator and uncover transition - direction dependent physiological trends.

The purpose of this work is to determine if center frequency alone, or in conjunction with traditional spectral features, improves a classifier's ability to correctly identify workload level. Additionally, this study serves to further investigate the evidence of physiological hysteresis when a human operator is suddenly subjected to an immediately changing workload.

It may be reasonable to assume that center frequency may not independently contribute to the improvement of classification accuracy when combined with traditional spectral features as it is derived from the spectral features. However, it may be possible that alone, center frequency calculated across all frequency bands will contain enough discriminating information between the two classes to accurately classify cognitive workload. Additionally, center frequency calculated across the alpha and theta band may not contain enough information for the accurate classification of workload. Finally, we would suspect that there may be an asymmetry in both the amount of lag as well as signal response, depending on the direction of the transition. Presumably, there will be a greater lag in response when the transition is from hard task difficulty to easy, as well as a greater decrease in signal. As a result, there may be higher classification accuracy when the transition is from hard task difficulty to easy due to a distinct separation of the signal between classes.

2. Background

a. Classic physiological trends relating to workload

Electroencephalography (EEG) has been employed in many studies to explore changes in cognitive workload during a variety of tasks including memory, pilot simulations and operational pilot flight (Berka, et al., 2007), (Brookings, Wilson, & Swain, 1996), (Wilson, Swain, & Ullsperger, 1999), (Gevins & Smith, 2005), (Jung, Makeig, Stensmo, & Sejnowski, 1997), (Kohlmorgen, et al., 2007). The recorded signal potential is representative of the superposition of field potentials produced by many simultaneously active neuronal components. Because EEG normally includes a composite of signals that make up the total signal from 0.5 to 100Hz, many analyses are conducted following the extraction of the Fourier components which may be further broken into the traditional frequency bands (Delta [0-3 Hz], Theta [4-7 Hz], Alpha [8-12 Hz], Beta [13-30 Hz] and Gamma [31-42 Hz]). These bands are highly dependent on the degree of activity in the cerebral cortex and represent the overall excitation of the brain resulting from functions in the brainstem reticular activating system (RAS). In general, as the frequency of the EEG signal increases, so too does the degree of cerebral activation (Table 1).

Table 1 Activity levels associated with each EEG frequency band.

Activity Levels Associated With Each EEG Frequency Band				
Delta	*Theta*	*Alpha*	*Beta*	*Gamma*
Deeply sleeping subjects, coma patients	Emotionally distressed subjects	Awake, relaxed subjects	Intense mental demand	

Initial studies consistently found that midline frontal Theta and parietal Alpha are indicators of changing workload levels (Grundel & Wilson, 1992), (Gevins, et al., 1979), (Wilson & Eggemeier, 1991), (Hankins & Wilson, 1998), (Trejo, et al., 2007) (Christensen, Estepp, Wilson, & Russell, 2012) (Fournier, Wilson, & Swain, 1999). In an investigation on location specific effects, Gevins et al found that frontal midline Theta increased as the memory load and comparison task difficulty increased. Conversely, parietocentral and posterior occipitoparietal Alpha tended to decrease with an increased spatial memory load. Practice of the task resulted in an increase in both the Theta and Alpha signals, likely as a result of a longer focus period and decreased requirements of cortical resources due to increased skill level, respectively (Gevins A. , Smith, McEvoy, & Yu, 1997) (Fairclough, Venables, & Tattersall, 2005).

In addition, increased low frequency signals in the Delta band located at central and parietal areas as well as increased levels of Beta signal may be correlated with increased workload during complex flight activities (Wilson G. , 2002). In tasks which last a few hours, subject performance tended to decrease over time due to fatigue while both Theta and Alpha increased and Beta decreased (Wilson & Eggemeier, 1991). Additionally, further testing has revealed workload detection potential in the higher frequency bands of Beta and Gamma (Laine, Bauer, Lanning, Russell, & Wilson, 2002). Although several studies have found that laterally located higher frequency EEG signals are heavily contaminated by facial muscle artifacts, it appears that the contaminated signal is still useful for cognitive workload classification (Monnin & Estepp, 2011).

b. Hysteresis

Hysteresis is broadly defined as the lag in a variable property of a system with respect to the effect producing it as this effect varies and can also be related to the dependence of a system on its past states. Performance hysteresis is the effect that workload transition type has on subject performance and is usually considered applicable when a system responds differently to identical inputs depending on the direction the system is driven (Farrell, 1999), (Helton, Shaw, J., Dember, & Hancock, 2004). Specifically, although sudden shifts in workload either from hard task difficulty to easy or easy difficulty to hard tended to cause a performance decrement, performance during increasing workload tends to increase until high difficulty workload is reached, at which point performance decreases (Cox - Fuenzalida, Beeler, & Sohl, 2006). Alternatively, performance tends to immediately decrease after a transition from high workload to low workload. A subject is most susceptible to the hysteresis effect after the changing workload level exceeds their capability to process information. Additionally, there is often a performance decrement when the subject is expecting a high input rate (difficult workload) and fails to notice the workload level has dropped (low workload). Further studies have demonstrated that, regardless of the transition type, the effect of hysteresis was more immediately significant when the transition between two workload levels was abrupt, rather than gradual, however gradual transitions result in longer term performance decrement (Moroney, Warm, & Dember, 1995).

In conjunction with performance effects, physiological signals may display a lag, or hysteresis effect, following a transition between two workload levels. This lag may occur for at least three reasons. First, if the transition is sudden, participants may not immediately notice or respond to the change in task difficulty, creating some lag between task difficulty change and cognitive workload impact. Additionally, there may be asymmetry in an operator's response time based upon the direction of the task difficulty transition (Cumming & Croft, 1973). Similarly, human operators may anticipate a transition in task difficulty resulting in little to no physiological signal lag. Lag in physiological systems may also arise as a function of the natural underlying control systems. For example, cardiac output rises and falls with changing metabolic demand, but the decline in heart rate is slow as compared to the rise (Wilmore & Costill, 2004). Finally, physiological systems are more or less remote from the neural activity associated with workload. We would expect that heart rate changes would lag more than EEG changes, as EEG is more directly measuring neural activity. In turn, different EEG measures may exhibit variable lag depending on the nature of the brain activity producing that signal.

c. Machine Learning

Machine learning algorithms are useful in EEG workload analysis because of their ability to mine vast amounts of data and differentiate the signal from noise, with high precision (Lemm, Blankertz, Dickhaus, & Muller, 2011), (Ling, Goins, Ntuen, & Li, 2001). A classifier is, in essence, the regression analysis of a discrete data set where the input features are created from a set of EEG data and the dependent output is a class level to which a particular window of data belongs (Pereira, Mitchell, & Botvinick, 2009) (Wilson G. , 1999) (Wilson & Fisher, 1995). In supervised learning approaches to classification, the classifier must first learn the classification parameters from a set of training data, where the learned classifier is a model of the features relative to the training data. The trained classifier is then tested on a new set of data (test data) and it determines the class level for the test data. The training data is assumed to be independent from the testing data though they may originate from the same distribution of data. One of the classifier's outputs is an accuracy which is the fraction of data in the test set which was correctly labeled:

$$(1)\ \frac{\sum_{i=1}^{n\ test} I(f(x_i),y)}{n_{test}}$$

Where $I(f(x_i),y) = 1$ if the class label was predicted correctly for that window, and $I(f(x_i),y) = 0$ if the class label is incorrectly predicted (Pereira, Mitchell, & Botvinick, 2009).

While there are many different classification methods available, this study focuses on the use of supervised linear and non – linear support vector machine (SVM) classifiers due to the commonality of this approach and success in a broad range of classification problems. An SVM classifier is a discriminative model that directly learns to differentiate between two classes by creating a maximally separating decision boundary from a set of labeled training data (Pereira, Mitchell, & Botvinick, 2009). Linear SVM is a somewhat simpler method, best used when there are a large number of input features relative to the number of data points. Non – linear SVM is advantageous when the number of data points far exceeds the number of input features, and can provide a more exact regression fit for truly non – linear data than its linear counterpart. In linear classifiers, each feature affects the prediction by its weight without interaction with other features (Pereira, Mitchell, & Botvinick, 2009). Non – linear classifiers are kernel based, where the kernel function replaces the linear dot product in the input space, thus allowing for a non – linear decision surface primarily driven by the interactions between non –

linear functions and features. If the test set used for deriving an accuracy estimate is truly independent of the training set, the classification accuracy output is an unbiased estimation of the true accuracy. True accuracy is the probability that the classifier will correctly label a new exemplar drawn at random from the same distribution that the training data came from, or the accuracy that would be obtained with an infinite number of data points in the test set (Pereira, Mitchell, & Botvinick, 2009). Thus, the precision of the classifier is based on the number of exemplars available for the test set, more exemplars increases precision. A statistically significant classification result is one where we can reject the null hypothesis that there is no information about the variable of interest in the data from which it is being predicted. This is how we will show the validity of the empirical tests.

d. Feature Selection

One advantage to using EEG for determination of cognitive workload level is that EEG provides fine, temporal resolution of activity with minimal invasiveness (Laine, Bauer, Lanning, Russell, & Wilson, 2002). This often results in the generation of a large amount of data, as well as many input features. Appropriate feature selection is vital in the success of a classifier. Feature extraction is the processing of raw data into sets of measures that quantify a group of states for classification (Russell & Gustafason, 2001). To reduce the effects of noise within training data, a large amount of data relative to the number of features for training the classifier is ideal. Additionally, using a large amount of data in the test set helps to increase the power of the test for significance of the resulting accuracy. Although averaging the training data may eliminate noise and smooth the signal, natural meaningful variability within a data set may also be reduced or eliminated. Another concern for feature and data selection is to choose a data set such that the number of data points in each class is equal to avoid any biasing of classes, which might affect the classifier accuracy.

Because a large number of features may be derived from any particular data set, especially in EEG data sets, it is often advantageous to select only the most interesting or useful features to increase the ratio of data points to features. Having approximately the same number of features as data points results in a falsely high classification accuracy (over fitting) as data points may be uniquely identified by their input features. For this study, the number of total possible features was significantly less than the number of data points used to train and test the classifier, thus reducing noise in the classification model and improving the overall classification accuracy of external validation data.

e. Center frequency

Center frequency is a fairly common measure used in signal processing applications to characterize a power spectrum. It is, by definition, the centroid or center of mass of a particular spectrum and may be calculated as the weighted mean of the signal frequencies determined from the power spectrum (Peeters, 2004). In application, center frequency relating to digital and audio signal processing represents the impression of sound brightness and is used as a direct measure of musical tone quality (Gordon & Grey, 1978). Similar relationships may be found for signals of optical light waves (Massar, Fickus, Bryan, Petkie, & Terzuoli, 2010), wireless transmitters, reflection seismology (Barnes, 1993) and other electronic devices.

Center frequency has been previously used for the quantitative analysis of neuromuscular activity during physical workload. After collection of a raw

Electromyogram (EMG) signal, processing begins with a differential amplifier, followed by a band-pass filter and full wave rectification before being analyzed by a Fast Fourier transform (FFT). The signal is then further processed to produce average power from the power spectral density (PSD), root mean square amplitude and center frequency. The center frequency is the frequency that divides the power spectrum into equal upper and lower portions. Previous studies have been conducted with EMG signals during static work where the center frequency as a function of fractional isometric force impulses remains constant with increasing isometric force impulses (Phillips, 2000). As muscles fatigue, the center frequency tends to decrease as a result of the slowing of the conduction velocity of motor unit action potentials on the sarcolema (Petrofsky & Phillips, 1985). Additionally, increasing muscle temperature tends to increase the center frequency due to an increase in conduction velocity of the muscular motor unit action potential. However, this trend is not enough to offset the effect of center frequency reduction in a sustained isometric contraction to the point of fatigue.

EEG signals display a high level of noise which can negatively affect the ability of the classifier to categorize the underlying signal and may lead to incorrect classification levels. To date, many different methods have been attempted to sift through the noise in the cortical signals (Cannon, Krokhmal, Chen, & Murphey, 2011) (Crossen, 2011) (Christensen, Estepp, Wilson, & Russell, 2012). There is some evidence that including center frequency in the EEG feature set may improve classification accuracy as center frequency provides a noise – robust estimate of how the dominant frequency is changing over time (Sun & Zhang, 2005), (Massar, Fickus, Bryan, Petkie, & Terzuoli, 2010), (Ngoc Le, Ambikairajah, Epps, Sethu, & Choi, 2011). Some research has been performed which attempts to use center frequency as a stress classifier in EEG recordings (Sulaiman, Taib, Aris, Hamid, Lias, & Murat, 2010). Sulaiman et al attempted to identify stress unique features from spectral centroid and cortical asymmetry where the subject has eyes closed and the subject has eyes open, performing an IQ test. When the subject was relaxed, the spectral centroid values were lowest for the Theta and Beta frequency bands on the right side of the brain, mid – range for the Delta, Theta and Beta frequency bands on the left side of the brain and highest for the Delta band on the right side as well as the Alpha band on the left and right side (Sulaiman, Taib, Aris, Hamid, Lias, & Murat, 2010). They did not find a clear pattern of differences when the subjects were engaged.

3. Objectives of the Study

The three objectives of this study are as follows:

(1) Examine center frequency for improved classification accuracy when used in conjunction with traditional spectral features or alone.

(2) Determine whether center frequency, calculated across the alpha and theta band performs significantly better than it does when calculated across all frequency bands.

(3) Evaluate asymmetry in classification accuracy between transition types hard to easy and easy to hard.

To further explore the three main questions, we will be employing supervised machine learning techniques to compare classification accuracy for various feature sets. The contents of the feature sets will address each of the different focuses of this study.

4. Materials and Methods

a. Multi-Attribute Task Battery Simulation Task

After obtaining IRB approval for the study, ten participants (six male, four female) ranging in age from 18 to 40 years, after providing signed, informed consent, were trained as operators in a Multi-Attribute Task Battery (MATB) simulation task originally developed by NASA (Comstock & Arnegard, 1992) and customized to run in MATLAB v.7.11.0 (R2010b). In this task, subjects were instructed to monitor various components of the software simulation which are analogous to those that a pilot or flight crew member might encounter. The MATB tasks include monitoring, tracking, communication and resource allocation components, all of which occur simultaneously in a continually changing environment. The demands of each task were individually controlled, such that the overall difficulty would be representative of either easy workload or hard workload. Because the ability of individuals to perform in the MATB environment may vary dramatically between subjects, each participant was trained on the task over several two hour sessions and a performance asymptote was individually determined and considered the high workload level for that subject. The same "low workload" level was used for each of the subjects. Low workload was a sufficiently easy level of effort which would be analogous to flying on "auto – pilot" and hard workload chosen such that the subject would not be able to correctly perform all tasks.

The subjects returned on a separate day for data collection and physiological and performance measures were recorded while the subject engaged with the MATB environment for a total of 24 six minute continuous runs (6 easy, 6 hard, 6 easy transitioning to hard and 6 hard transitioning to easy).

For the easy runs, the subjects interacted with the MATB environment at a broadly defined easy level for a full six minutes without a transition to hard. For hard runs, subjects performed at an individually determined hard workload level for a full six minutes without a transition to easy. The transition between workload levels is the focus of this study, more specifically, the transition from easy to hard as well as the transition from hard to easy. For these runs, participants were exposed to six runs, each containing three minutes of easy workload immediately followed by three minutes of hard workload, then six runs, each containing three minutes of hard workload immediately followed by three minutes of easy workload. Two minute breaks were allocated between runs for rest.

b. Physiological Data

A total of seven channels were recorded continuously using a Cleveland Medical Devices BioRadio 110 telemetry unit linked with the New Workload Assessment Monitor (NuWAM). NuWAM includes algorithms developed by the Air Force Research Laboratory and has the capability of removing eye blink and movement artifacts in the EEG signals via the use of an adaptive filter algorithm (Krizo, Wilson, & Russell, 2005). Five EEG channels (FZ, F7, PZ, T5, O2) as well as one channel for vertical electrooculogram (VEOG) and one channel for horizontal electrooculogram (HEOG) were collected. EEG electrodes were placed on the scalp in accordance to the 10-20 System for Electrode Placement and held in place using a traditional electrode cap. VEOG and HEOG were recorded for artifact correction purposes.

Within NuWAM, the EEG data was segmented and the power spectral density (PSD) was calculated for each segment with a one second resolution to determine features that could be used to assess the effects of changing workload on continuous brain activity. More specifically, the data was originally sampled at 200 Hz, and then was band pass filtered from 0.5 Hz to 100 Hz with 60 Hz line noise notched out**.** Vertical and horizontal eye artifacts were corrected using an embedded linear-regression technique where VEOG and HEOG were used as representative artifact signals in the regression. Transformation into the frequency domain using the Nyquist frequency was accomplished using a moving window (10 seconds with 9 second overlap) and the power spectrum was generated and saved for each channel individually.

c. Feature Selection

The power spectral density (PSD) for each channel was further divided into traditional spectral frequency bands (Delta [0-3 Hz], Theta [4-7 Hz], Alpha [8-12 Hz], Beta [13-30 Hz] and Gamma [31-42 Hz]). Feature extraction and workload classification was performed in MATLAB v.7.11.0 (R2010b). Modifications to existing code allowed for specific features to be extracted and utilized. The clinical frequency bands were extracted for each channel of data at a 1 Hz interval and stored for further analysis. Concurrently, the band powers are utilized to calculate the center frequency across all frequency bands (Delta – Gamma) as well as across only the Theta and Alpha band. Published research on EEG indices of workload has shown that the Alpha and Theta bands exhibit workload-dependent changes that are likely reflective of changes in cortical activity (Gevins, et al., 1979). Changes in higher frequency signals, while useful in workload detection, are more likely to reflect the influence of muscular artifacts.

The center frequency was also considered, and was calculated as the sum of the frequency times the PSD divided by the sum of the PSD to produce the centroid of the spectrum, as shown below:

$$(2)\ CF = \frac{\sum F*PSD(f)*\Delta F}{\sum PSD(f)*\Delta F}$$

Where F represents the weighted frequency values and PSD (f) is the power spectral density over the frequency range for each window. Addition of center frequency calculated over all frequency bands and center frequency within Theta/Alpha band for each of the seven channels resulted in an additional 14 features for use in the classifier. To create a more normal Gaussian distribution within the data, the log mean power is taken for each of the band powers as well as the center frequency data. Often, recorded EEG data is skewed due to various artifacts. Log transformation helps to correct this effect and create a more appropriately normalized set of data for analysis.

The band power data was then windowed with a 10 second window with 9 seconds of overlap, resulting in a sample length of one second. Windowing creates a smoother signal by averaging over the window length, which can be useful in reducing outliers. This allows the data to be more reflective of the true brain signal which is slowly adapting without accentuating the noise component which may be affected by rapidly changing Electromyogram artifacts. Windowing also appears to improve classification accuracy for determining workload levels (Crossen, 2011). There are potential drawbacks to windowing, including a false correlation between windows and a false classification of artifacts but distinct separation prior to windowing of training and testing data for the classifier helps to overcome these concerns (Crossen, 2011).

Finally, the data is organized into a total of 51 features which includes five band power and two center frequency variables for each of the seven channels (Table 2).

Table 2 Comparison of feature sets - Shows contents of each feature set.

Comparison of Feature Sets			
Spectral	**Spectral+CF**	**CF All Bands**	**CF Alpha/Theta Band**
(7 Channels * 5 Frequency Bands) + VEOG + HEOG = 37 Features	Spectral + (7 Channels * 1 Center Frequency All Bands) = 44 Features	7 Channels * 1 Center Frequency All Bands = 7 Features	7 Channels * 1 Center Frequency Alpha/Theta Band = 7 Features

After features have been created for all the runs for each subject, training and testing vectors must be extracted for use in the classifier.

Figures 1 – 4 illustrate the effect a sudden transition has on the underlying signal being classified. Although there were five sites with five frequency bands total, T5 represents a site where EMG activity is likely a factor due to its location low on the scalp while FZ tends to be more representative of true cortical signal as it is located more towards the top of the scalp. Additionally, Theta traditionally demonstrates changes due to a sudden change in task difficulty, while Gamma shows changes more reflective of muscle tension. Both T5 and FZ Theta show a large and quick transition when the task difficulty changes from easy to hard with a minimal lag in signal; however for the hard to easy case, the transition is generally slower, but with much less noise (Figure 1 – 2). Having a significant amount of noise in the signal for the easy to hard transition direction may cause errors in classification because the two classes are not clearly distinct from one another, while a slow transition for the hard to easy case may create a lag in classification of the second class. T5 Gamma shows a somewhat similar trend where the signal exhibits a quick transition when the task difficulty changes from easy to hard with a minimal lag in signal; however again, for the hard to easy case, the transition is generally slower, but with much less noise (Figure 3). FZ Gamma is the least distinct of the four, showing overlap between both the classes (easy versus hard) and the transition direction (Figure 4).

Figure 1 Plot of power in the Theta band averaged across all subjects and all runs to show the effects of an immediate transition on T5 Theta. In this figure, and figures 2-4, the solid blue line represents the normalized log power of the signal during a run that transitioned from easy to hard task difficulty where the blue dotted line represents the corresponding positive and negative standard error of the mean. Similarly, the solid red line represents the normalized log power of the signal during a run that transitioned from hard to easy task difficulty with the dotted red line representing the positive and negative standard error of the mean. The dotted black line illustrates the exact moment the task difficulty transitioned.

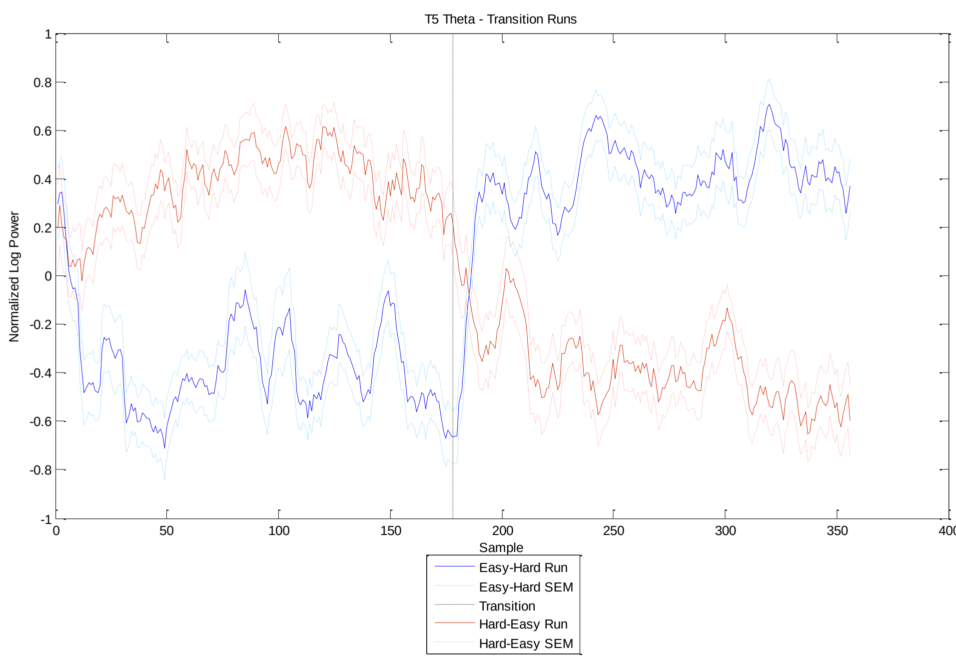


Figure 2 Plot of power in the Theta band averaged across all subjects and all runs to show the effects of an immediate transition on FZ Theta.

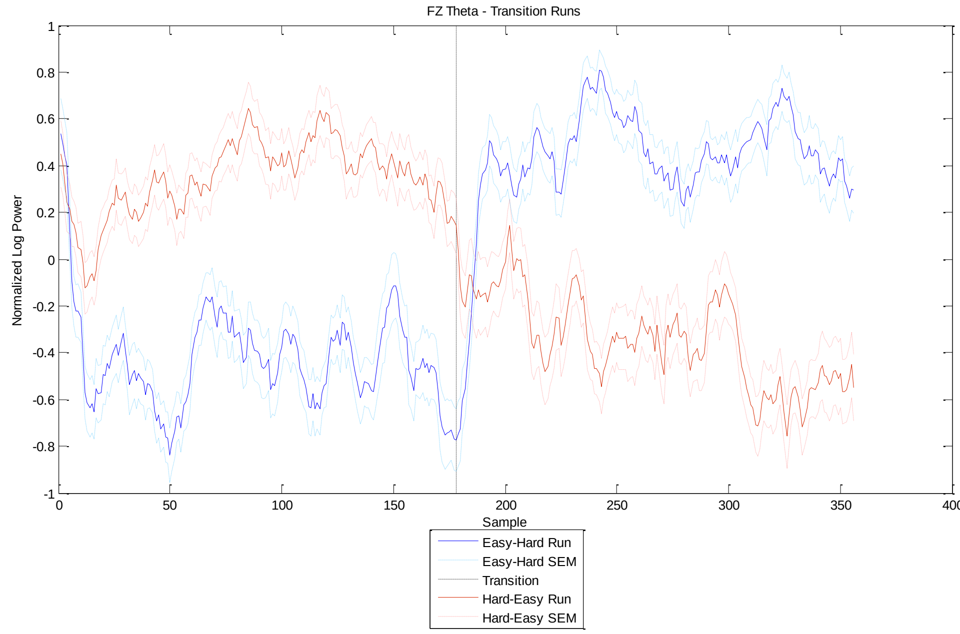


Figure 3 Plot of power in the Gamma band averaged across all subjects and all runs to show the effects of an immediate transition on T5 Gamma.

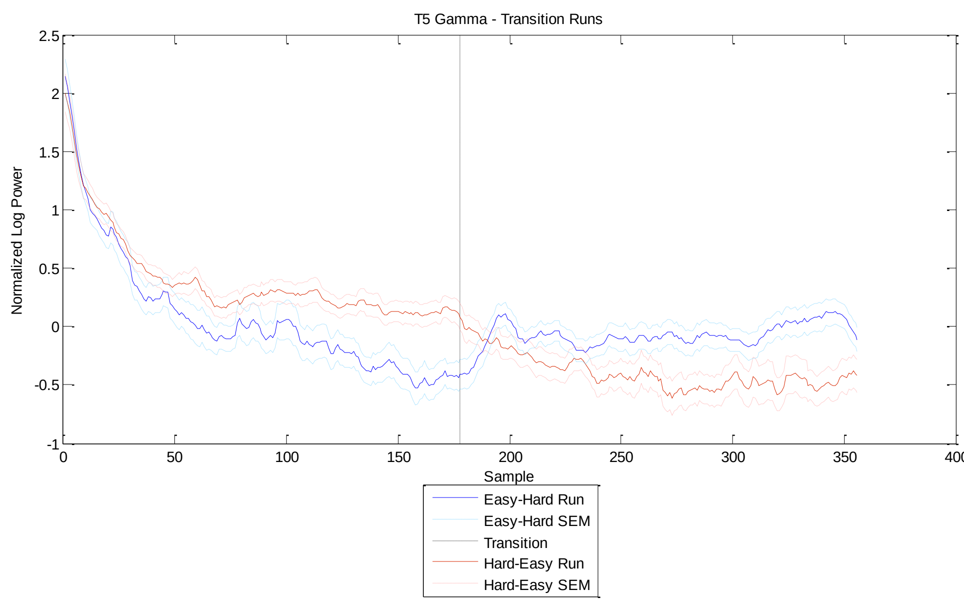
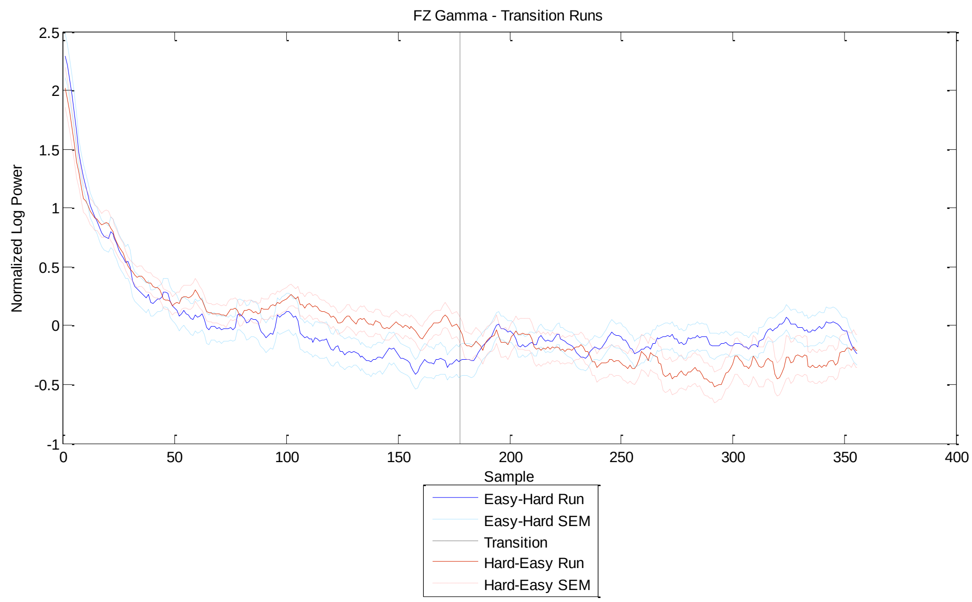


Figure 4 Plot of PSD averaged across all subjects and all runs to show the effects of an immediate transition on FZ Gamma.



Figures 5 - 8 illustrate the effect a sudden transition has on the center frequency calculated across all frequency bands as well as calculated across only the Alpha and Theta band. Center frequency calculated at T5 and FZ across the Alpha and Theta band and calculated at FZ across all frequency bands shows a sharp decline when the transition direction is easy to hard, and a gradual increase when the transition direction is reversed, with high levels of noise in both signals (Figures 5, 7 & 8). Calculated across all frequency bands at T5, center frequency does not differ significantly whether the transition is from easy to hard or hard to easy (Figure 6).

Figure 5 Plot of center frequency calculated across the alpha and theta band averaged across all subjects and all runs to show the effects of an immediate transition on T5. In this figure, and figures 6 – 8, the solid blue line represents the normalized log center frequency of the signal during a run that transitioned from easy to hard task difficulty where the blue dotted line represents the corresponding positive and negative standard error of the mean. Similarly, the solid red line represents the normalized log center frequency of the signal during a run that transitioned from hard to easy task difficulty with the dotted red line representing the positive and negative standard error of the mean. The dotted black line illustrates the exact moment the task difficulty transitioned.

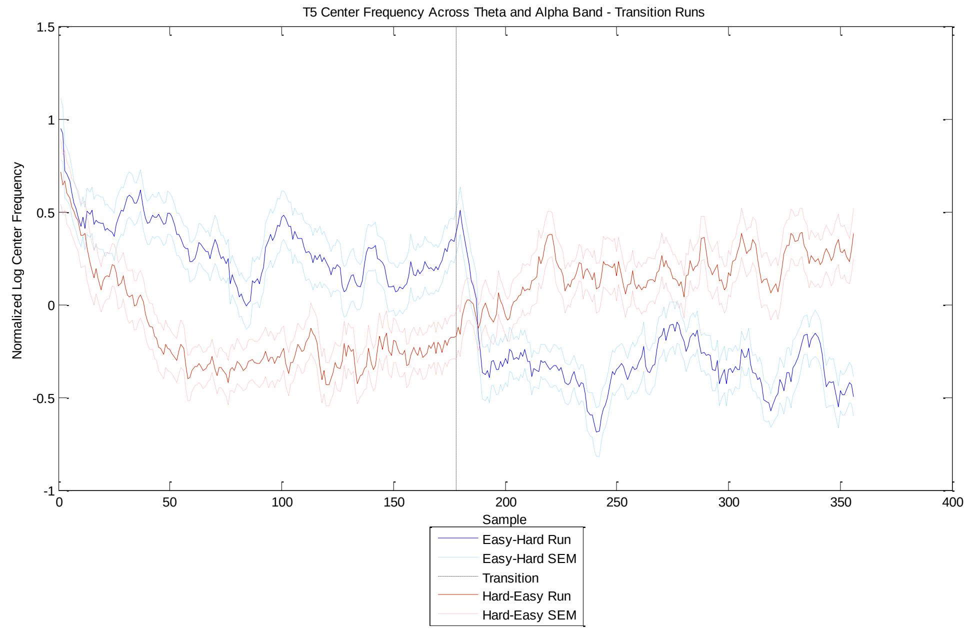


Figure 6 Plot of center frequency calculated across all frequency bands averaged across all subjects and all runs to show the effects of an immediate transition on T5.

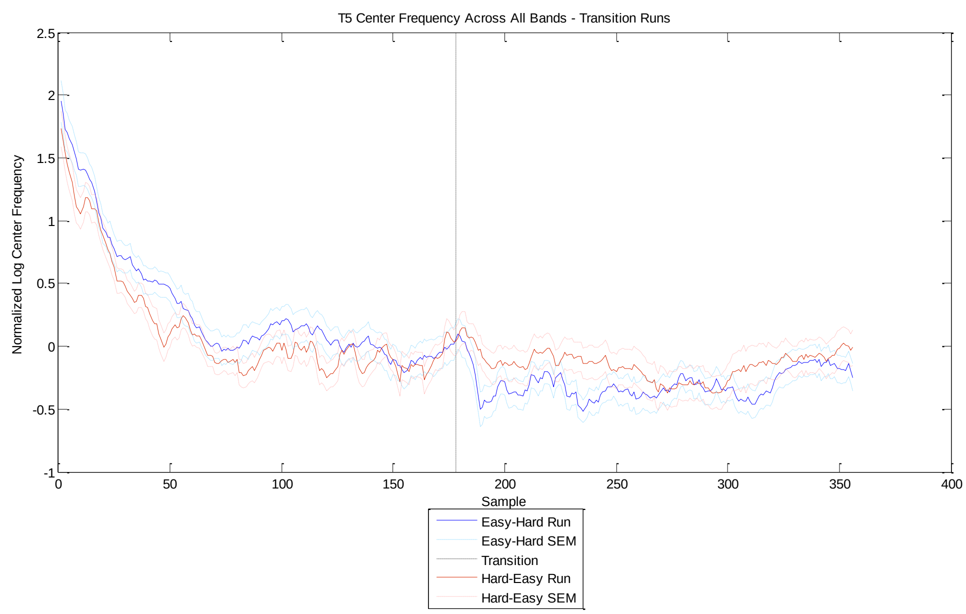
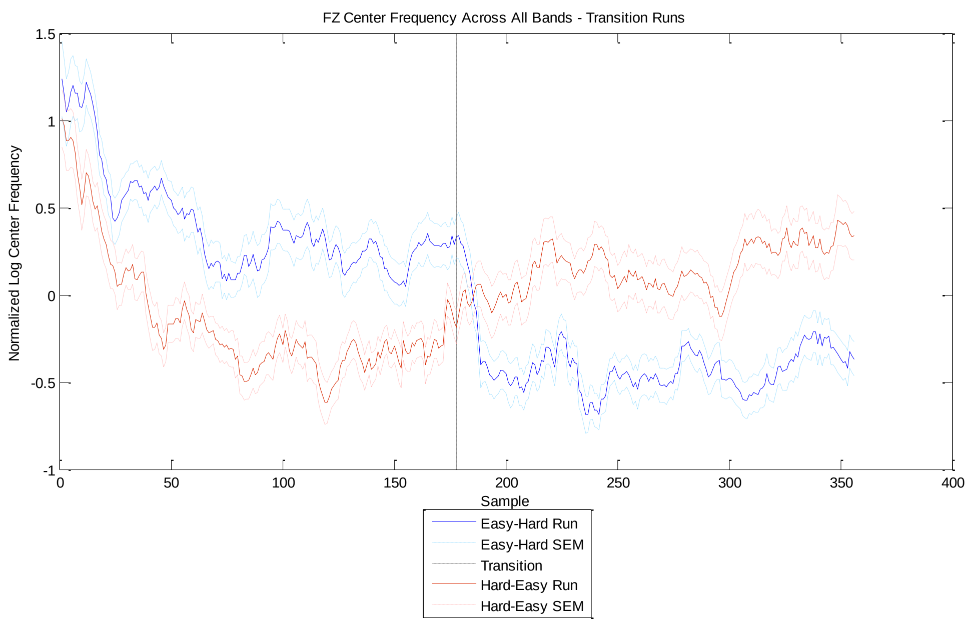


Figure 7 Plot of center frequency calculated across the alpha and theta band averaged across all subjects and all runs to show the effects of an immediate transition on FZ.

Figure 8 Plot of center frequency calculated across all frequency bands averaged across all subjects and all runs to show the effects of an immediate transition on FZ.



d. Support Vector Machine Analysis

A support vector machine (SVM) classifier is used to differentiate between workload levels where overall classification is nonlinear in input space. SVMs are ideal for brain imaging data applications due to their high generalization when the number of samples in significantly lower than the number of inputs (Pereira, Mitchell, & Botvinick, 2009). For the data in this study, the SVM was trained using a feature set composed of VEOG, HEOG and EEG data with an equal number of exemplars from the easy and hard workload levels. The training and testing sets were created from all data available for each subject and included all six runs of each transition type. The training set was composed of the six easy and six hard straight runs where no transition was present (approximately sixty six percent of the total data). The testing sets were composed of either six hard to easy transitions or six easy to hard transitions (approximately thirty three percent of the data) (Table 3).

Creating the testing and training sets in this fashion provides a large amount of sample data (exemplars), relative to the number of input features, improving the statistical relevance of the classifier output. This does, however, assume that the signal distributions associated with low and high workload are consistent from one run to the next within a transition type, and from one transition type to another. The first assumption is reasonable, as Christensen, Estepp, Wilson, and Russell (2012) demonstrated that good classification accuracy is possible as long as the training set is representative of the span of time included in the test set. The second assumption is one of the key questions of this study; if signals following a sudden transition are significantly different from those in a consistent workload run, we would expect to see degraded classification accuracy.

Table 3 Number of exemplars in training and testing vectors. Number of exemplars training is a compilation of all the low and high workload straight runs for a total of approximately 4300 exemplars. Number of exemplars testing is dependent on the transition type, so there are only about 2100 exemplars available for the classifier. The maximum size of input feature set for the SVM is 44 which is the case for spectral features plus center frequency across all bands. Additional sets include spectral features alone (37), center frequency across all bands (7) and center frequency across the Alpha and Theta band (7).

	Single 6 min. Run		All 6 min. Runs (6)			
Run Type	**# Exemplars Easy**	**# Exemplars Hard**	**# Exemplars Easy**	**# Exemplars Hard**	**# Exemplars Training Set**	**# Exemplars Testing Set**
Easy	360	0	2160	0	4320	----------------------
Hard	0	360	0	2160		----------------------
Easy to Hard	180	180	1080	1080	-----------------------	2160
Hard to Easy	180	180	1080	1080	-----------------------	2160

In order to assess the effects of center frequency on workload classification accuracy, the SVM was trained and evaluated using four combinations of features. This evaluation was conducted separately for the two (low to high and high to low) transition types in the test set. The first of the four feature sets contained all 37 of the traditional spectral features (Table 2). The second feature set contained the 37 spectral features plus the center frequency calculated across all frequency bands for each of the seven electrode sites for a total of 44 features. The third feature set contained only the center frequency calculated across all bands for the seven electrode sites for a total of seven features and the final set contained only the center frequency calculated across the Theta and Alpha frequency bands for a total of seven features. Theta and Alpha were chosen for the final feature set based on the previous research that suggests a possible correlation between task difficulty and changes in Theta and Alpha (Gevins, et al., 1979). The training and testing data sets for each SVM were normalized via Z – transform separately for each participant and feature by extracting means and standard deviations from the non-transition runs, and applying them to all runs:

$$(3)\ \frac{X-\mu}{\sigma}=z$$

Where μ is the mean of the population and σ represents the standard deviation of the population. The normalization process of log center frequency resulted in a dimensionless ratio centered on zero. As such, a shift positive or negative represents the underlying shift in signal to the right or left, respectively. As task difficulty increases, Theta tends to increase which causes a mean center frequency shift to the left, resulting in a decrease of the normalized signal.

The SVM classifier was then run eight separate times for each subject. Each run of the classifier correlated to a combination of the four feature sets with a transition direction (Table 4).

Table 4 Specification of SVM Cases - Describes the combination of feature set and transition direction for all eight runs of the SVM.

Specification of SVM Cases								
Feature Set	Spectral		Spectral+CF		CF All Bands		CF Alpha/Theta Band	
Transition Dir.	Low to High	High to Low	Low to High	High to Low	Low to High	High to Low	Low to High	High to Low

e. Experimental Design and Analysis

The transition type (easy to hard or hard to easy) is specified for each feature set type (spectral only, spectral plus center frequency all bands, center frequency all bands only, center frequency Alpha/Theta band) for a total of eight unique feature sets which were run through separate linear and non-linear SVM classifiers (Table 5). The classification accuracy (CA) output from the SVM classifier represents the percent of data in the test set which was correctly labeled as either high workload or low workload. For statistical analysis, a 4 (feature set type) x 2 (transition types) repeated – measures two – way Analysis of Variance (ANOVA) was conducted on the classification accuracy output from the linear and non – linear SVM separately using the Huynh-Feldt correction for any violations of sphericity. The data is assumed to be approximately normal, the variance homogeneous, and the conditions of sphericity are satisfied.

The interaction effect present in the linear SVM classification output was further investigated with several one-way repeated - measures ANOVAs which compared the two transition types for each feature set individually.

5. Results

Of the two main effects being tested (feature set and transition type), feature set showed significant differences with regards to classification accuracy for both the linear and non – linear SVM classifiers, $F_{Linear}(3, 27)= 10.371$, $F_{Non\text{-}Linear}(3, 27)= 25.984$, $p < 0.05$ (Tables 6-7). Interestingly, in the linear SVM output only, there was a significant interaction effect present between the two factors, feature set type and transition type, $F(3, 27) = 6.05$, $p = 0.0027$. There is no significant interaction between the transition type and feature sets for the non – linear case, indicating a lack of notable asymmetry between transition types.

Table 5 Overall classification accuracy results show average classification accuracy and standard deviation for each of the four feature set types and the corresponding transition direction type. Calculations are done using linear SVM and non - linear SVM classifiers.

Comparison of Classification Accuracies (CA) for all Feature Sets (FS) and SVM Types								
	Spectral FS Avg. CA		Spectral+CF Avg. CA		CF All Bands Avg. CA		CF Alpha/Theta Band Avg. CA	
	E to H	H to E	E to H	H to E	E to H	H to E	E to H	H to E
Linear SVM	77.15 ± 9.40	77.67 ± 15.18	76.93 ± 9.63	78.57 ± 14.36	68.36 ± 10.08	78.08 ± 8.03	68.24 ± 8.31	66.58 ± 8.48
Non-Linear SVM	76.97 ± 8.59	77.58 ± 14.78	77.44 ± 8.53	79.33 ± 14.28	66.40 ± 9.69	73.72 ± 8.89	63.33 ± 7.16	62.94 ± 6.79

Table 6 Two - Way Repeated Measures ANOVA results indicate feature set (F) to be a significant main effect with a significant interaction effect present between transition type (T) and feature set for the linear SVM.

Source	SS	df	MS	F	p
T	130.751	1	130.751	1.0699	0.328
F	**1.39E+03**	**3**	**464.554**	**10.371**	**1.03E-04**
T*F	**370.742**	**3**	**123.581**	**6.0531**	**0.0027**
T*Subj	1.10E+03	9	122.211		
F*Subj	1.21E+03	27	44.7958		
T*F*Subj	551.236	27	20.4161		

Table 7 Two - Way Repeated Measures ANOVA results indicate feature set (F) to be a significant main effect with no significant interaction effect present between transition type (T) and feature set for the non – linear SVM.

Source	SS	df	MS	F	p
T	111.011	1	111.011	1.0587	0.3304
F	**3.01E+03**	**3**	**1.00E+03**	**25.984**	**4.07E-08**
T*F	177.378	3	59.1261	2.7415	0.0627
T*Subj	9.44E+02	9	104.854		
F*Subj	1.04E+03	27	38.6434		
T*F*Subj	582.32	27	21.5674		

For the feature set types of spectral, spectral plus center frequency and center frequency Alpha/Theta, there was no significant difference in classification accuracy between the transition types, $F_{\text{CF Alpha/Theta}}\ (1, 9) = 0.63$, $F_{\text{Spectral}}\ (1, 9) = 0.25$, $F_{\text{Spectral plus CF}}\ (1, 9) = 0.02$, $p > 0.05$ (Tables 9-11). However, there was a significant difference between the transition types for the feature set center frequency all bands, $F\ (1, 9) = 9.37$, $p = 0.0135$ (Table 8).

Table 8 One - Way Repeated Measures ANOVA results indicate a significant difference in the linear SVM classification accuracy output (CA) between the two transition types for center frequency all bands.

Source	SS	df	MS	F	p
CA	**472.93**	**1**	**472.931**	**9.37**	**0.0135**
Subjects (matching)	1041.8	9	115.757	2.29	0.116
Error	454.17	9	50.463		
Total	1968.9	19			

Table 9 One - Way Repeated Measures ANOVA results indicate no significant difference in the linear SVM classification accuracy output (CA) between the two transition types for center frequency Alpha/Theta band.

Source	SS	df	MS	F	p
CA	13.8	1	13.802	0.63	0.4478
Subjects (matching)	1071.1	9	119.006	5.43	0.0095
Error	197.23	9	21.914		
Total	1282.1	19			

Table 10 One - Way Repeated Measures ANOVA results indicate no significant difference in the linear SVM classification accuracy output (CA) between the two transition types for spectral features.

Source	SS	df	MS	F	p
CA	13.36	1	13.364	0.25	0.6297
Subjects (matching)	2206.4	9	245.156	4.57	0.0168
Error	483.04	9	53.6713		
Total	2702.8	19			

Table 11 One - Way Repeated Measures ANOVA results indicate no significant difference in the linear SVM classification accuracy output (CA) between the two transition types for the spectral plus center frequency feature set.

Source	SS	df	MS	F	p
CA	1.4	1	1.397	0.02	0.8795
Subjects (matching)	2352.9	9	261.433	4.55	0.017
Error	516.7	9	57.411		
Total	2871	19			

The lack of significant differences between transition types for each feature set supports the evidence of no interaction between main effects for the non – linear SVM output. This is shown in Tables 12 – 15.

Table 12 One - Way Repeated Measures ANOVA results indicate no significant difference in the non - linear SVM classification accuracy output (CA) between the two transition types for center frequency all bands.

Source	SS	df	MS	F	p
CA	267.96	1	267.96	4.5	0.062
Subjects (matching)	1023.3	9	113.7	1.9	0.173
Error	533.4	9	59.267		
Total	1824.6	19			

Table 13 One - Way Repeated Measures ANOVA results indicate no significant difference in the non - linear SVM classification accuracy output (CA) between the two transition types for center frequency Alpha/Theta bands.

Source	SS	df	MS	F	p
CA	0.785	1	0.7848	0.1	0.788
Subjects (matching)	783.38	9	87.042	8.5	0.002
Error	92.061	9	10.229		
Total	876.23	19			

Table 14 One - Way Repeated Measures ANOVA results indicate no significant difference in the non - linear SVM classification accuracy output (CA) between the two transition types for spectral feature set.

Source	SS	df	MS	F	p
CA	1.89	1	1.893	0	0.858
Subjects (matching)	2125.9	9	236.21	4.2	0.022
Error	503.82	9	55.98		
Total	2631.6	19			

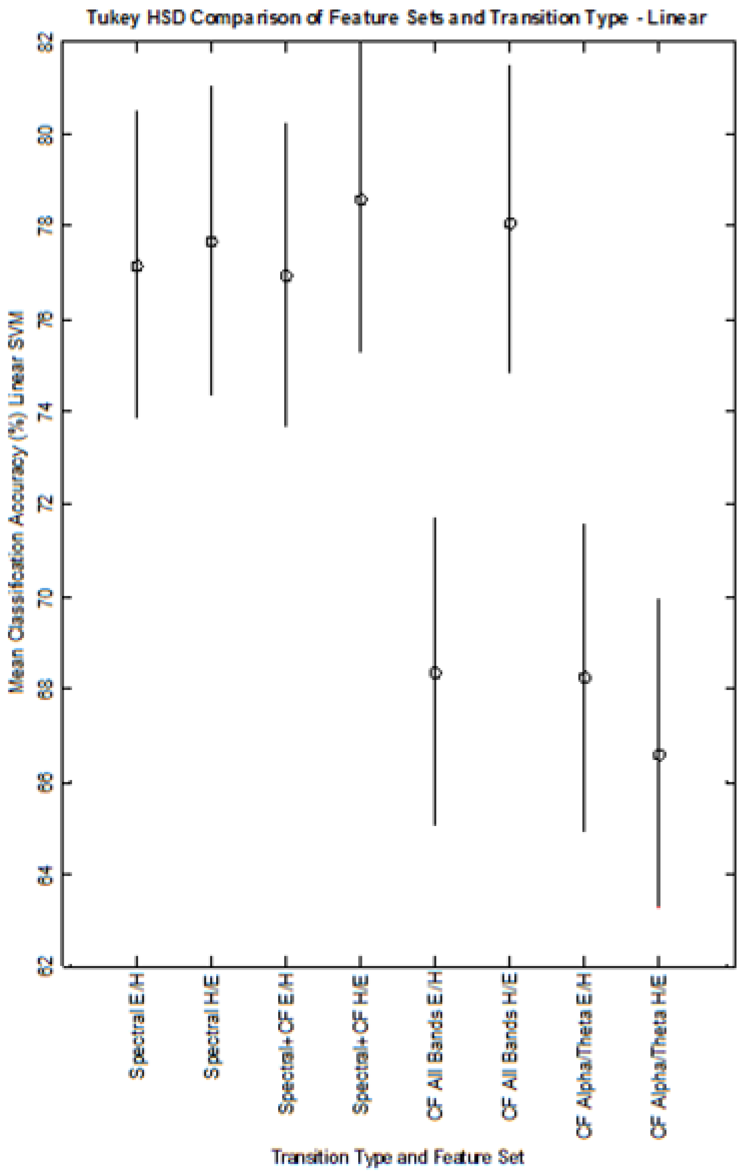
Table 15 One - Way Repeated Measures ANOVA results indicate no significant difference in the non - linear SVM classification accuracy output (CA) between the two transition types for spectral plus center frequency feature set.

Source	SS	df	MS	F	p
CA	17.75	1	17.747	0.4	0.542
Subjects (matching)	2093	9	232.56	5.3	0.011
Error	396.73	9	44.081		
Total	2507.5	19			

Tukey HSD (Honestly Significant Difference) post – hoc comparison of mean classification accuracy output from the linear SVM illustrates significant differences between various feature sets and transition directions (Figure 9). Generally, the spectral only and spectral plus center frequency feature sets ($M_{\text{Spectral E/H}}$ = 77.1462, $M_{\text{Spectral H/E}}$ = 77.6748, $M_{\text{Spectral+CF E/H}}$ = 76.9336, $M_{\text{Spectral+CF H/E}}$ = 78.5685, SD = 1.4288) classify significantly better than center frequency alone. Additionally, there does not appear to be a significant improvement in classification accuracy when center frequency is added to the spectral feature set, possibly due to the fact that center frequency is calculated from components already accounted for in the spectral feature set. Another explanation for this result is the presence of a ceiling effect. In previous results (Wilson & Russell, 2003), a classification accuracy of 80% is generally the maximum observed for realistic tasks. If 80% does indeed represent a ceiling on workload classification accuracy, there is not a great deal of improvement that is possible with the addition of center frequency to the spectral features as compared to spectral features alone. So, while center frequency may indeed be contributing to the classifier, there is not likely to be a significant additive effect visible.

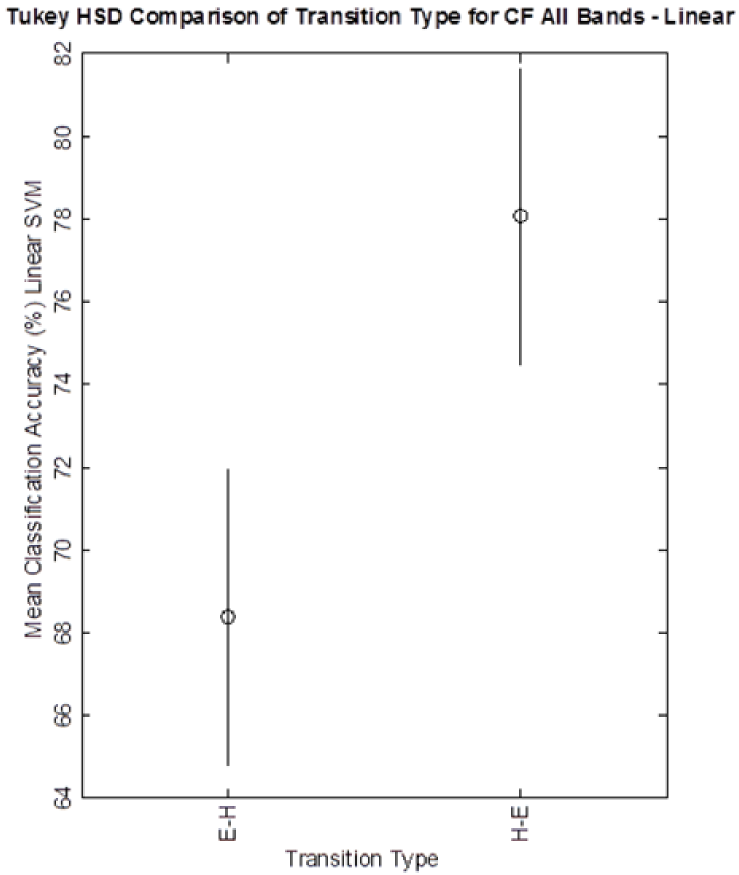
Notably, for the case where workload changes from hard to easy, the linear SVM is equally proficient in classifying workload level using only the center frequency feature set as it is using the full spectral feature set ($M_{CF\ All\ bands}$ = 78.0841, SD = 1.4288 versus $M_{Spectral\ H/E}$ = 77.6748, SD = 1.4288) (Figure 9). Within the linear classification output for center frequency all bands, there is a distinct asymmetry present with regards to the two transition types. The mean classification accuracy for the easy to hard transition condition is significantly poorer than that for hard to easy ($M_{CF\ All\ bands\ E/H}$ = 68.3585 versus $M_{CF\ All\ bands\ H/E}$ = 78.0841). Center frequency across only the Alpha and Theta bands does not appear to be a good single metric for the linear classifier based on mean classification accuracy, regardless of transition direction ($M_{CF\ A/T\ E/H}$ = 68.2440, $M_{CF\ A/T\ H/E}$ = 66.5825, SD = 1.4288) (Figure 9).

Figure 9 Tukey HSD comparison of mean classification accuracy outputs from the linear SVM indicate significant differences between the four feature sets and the two transition directions.



Additional Tukey HSD post – hoc comparison of mean classification accuracy output from the linear SVM supports previous findings of significant differences between the transition directions for the center frequency all bands feature set (Figure 10). The mean classification output for the transition type hard to easy (M = 78.0841, SD = 2.8829) was significantly higher than the mean classification output for the easy to hard transition (M = 68.3585, SD = 2.2464) (Figure 10).

Figure 10 Tukey HSD comparison of mean classification accuracy outputs from the linear SVM indicate a significant difference between the transition types Hard to Easy and Easy to Hard for the feature set center frequency all bands.



Similarly, Tukey HSD post – hoc comparison of mean classification accuracy output from the non - linear SVM illustrates significant differences between various feature sets and transition directions (Figure 11). Generally, the spectral only and spectral plus center frequency feature sets ($M_{Spectral\ E/H}$ = 76.9650, $M_{Spectral\ H/E}$ = 77.5803, $M_{Spectral+CF\ E/H}$ = 77.4417, $M_{Spectral+CF\ H/E}$ = 79.3257, SD = 1.4686) classify significantly better than center frequency across the Alpha/Theta bands and center frequency all bands for E/H ($M_{CF\ Alpha/Theta\ E/H}$ = 63.3383, $M_{CF\ Alpha/Theta\ H/E}$ = 62.9421, $M_{CF\ Allbands\ E/H}$ = 66.4029, SD = 1.4686). Initial comparison results indicate a significant difference between transition types in the non – linear classification output for the center frequency all bands feature set. The non – linear SVM classifier performed significantly more accurately when the workload level changes from hard to easy rather than easy to hard ($M_{\text{CF All Bands E/H}}$ = 66.4029, $M_{\text{CF All Bands H/E}}$ = 73.7236, SD = 1.4686). However, additional post-hoc comparison of mean non – linear classification output for the center frequency all bands feature set, indicated mean classification accuracy for the transition type hard to easy (M = 73.7236, SD = 2.4345) to be significantly higher than mean output for the easy to hard transition (M = 66.4029, SD = 2.4345) (Figure 12).

Figure 11 Tukey HSD comparison of mean classification accuracy outputs from the non - linear SVM indicate significant differences between the four feature sets and the two transition directions.

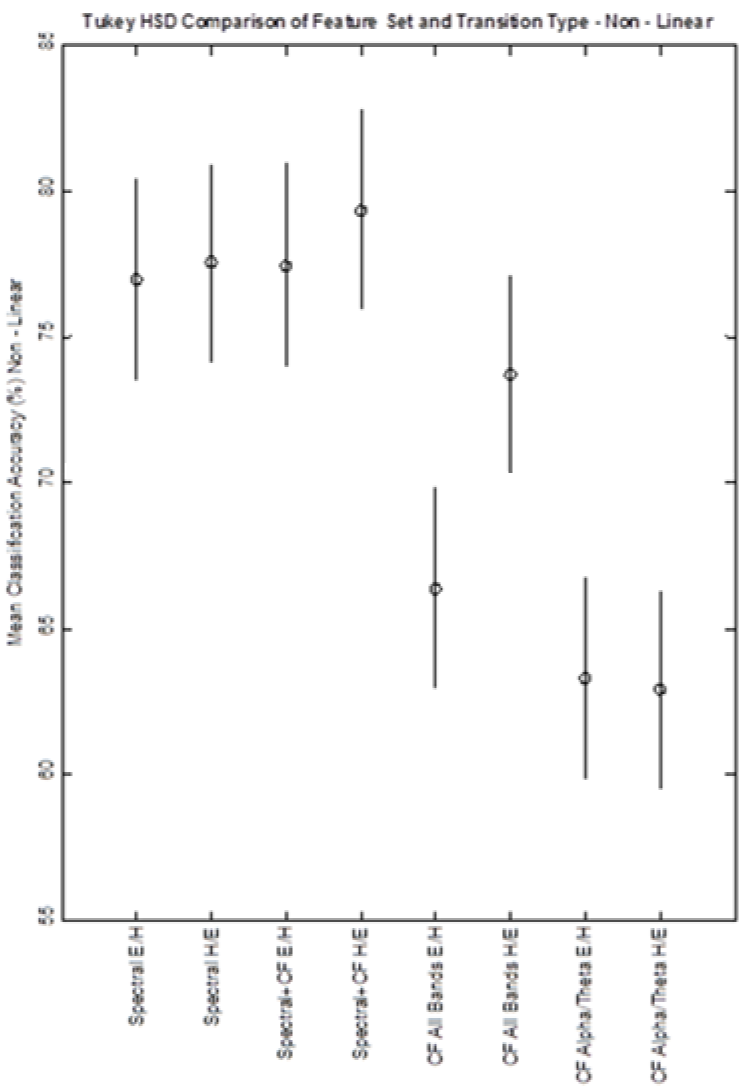
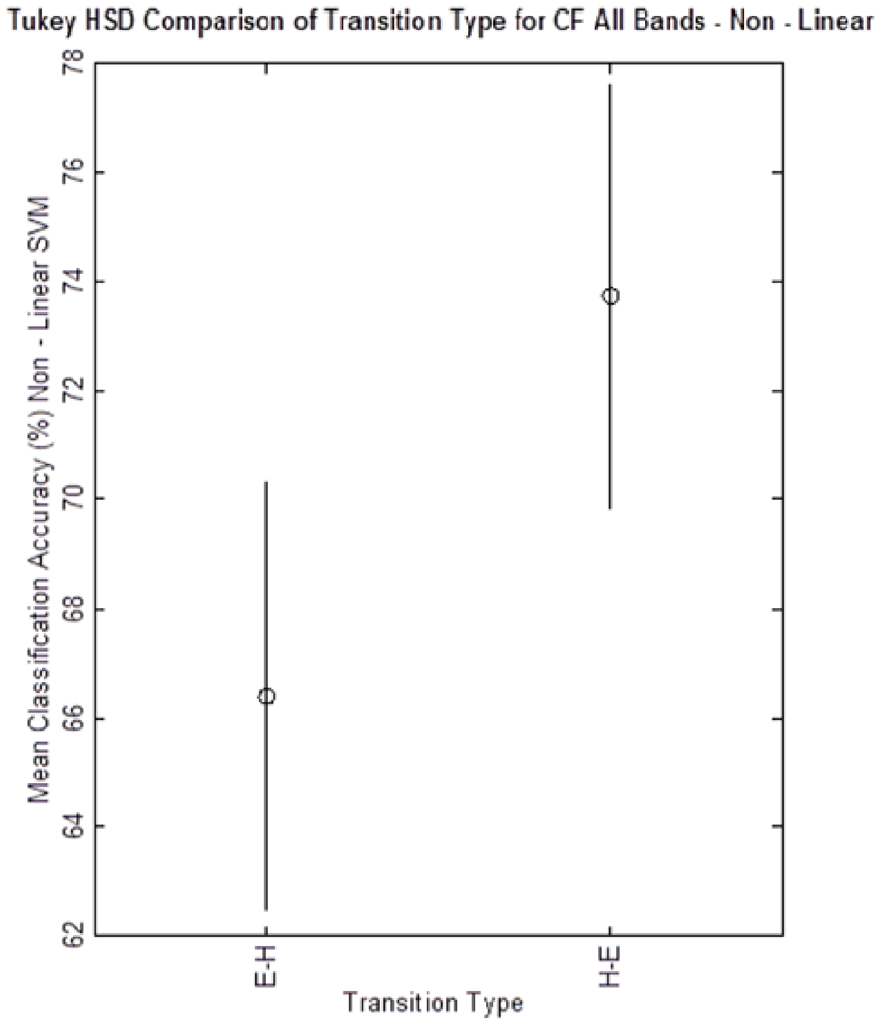


Figure 12 Tukey HSD comparison of mean classification accuracy outputs from the non - linear SVM indicate no significant difference between the transition types Hard to Easy and Easy to Hard for the feature set center frequency all bands.

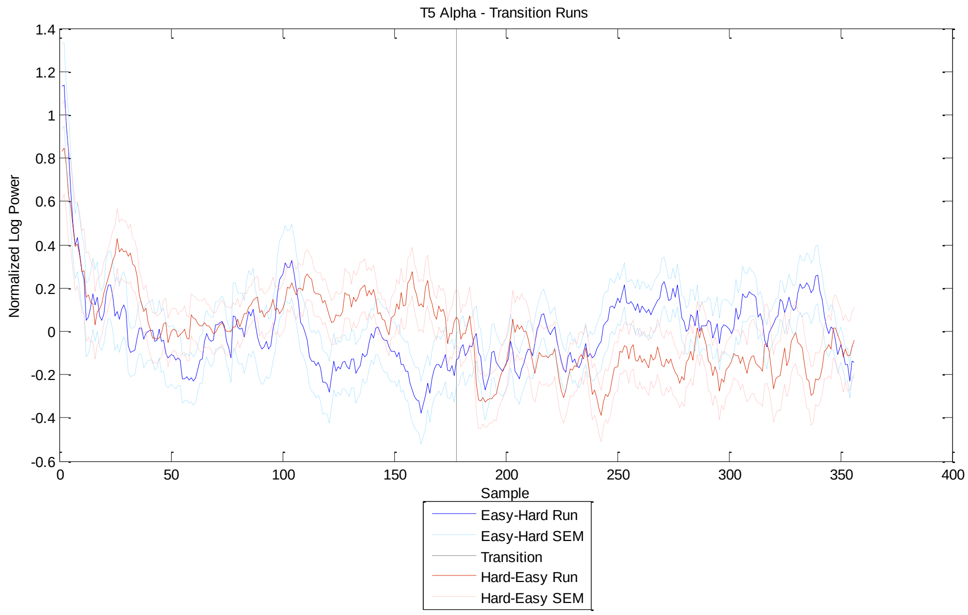


6. Discussion

This work set out to determine if center frequency alone, or in conjunction with traditional spectral features, improved a classifier's ability to correctly identify workload level. Additionally, this study further investigated evidence of physiological hysteresis when a human operator is suddenly subjected to an immediately changing workload. The three main observations from this study are as follows: (1) the confirmation of mean spectral power, spectral plus center frequency and center frequency all bands (high to low workload) data being the most robust feature sets for workload classification, (2) the observation of center frequency calculated across all frequency bands being an equally reliable metric for classifying workload when the level transitions from hard to easy and (3) asymmetry in classification accuracy between transition types hard to easy and easy to hard for the feature set center frequency across all bands. The comparison of classification accuracy between feature set types and transition directions was found using workload classification pre-processing methods and quantified by Tukey HSD mean comparison statistical analysis of the mean classification accuracy.

From the Tukey HSD analysis, the EEG - derived feature sets containing traditional spectral features and spectral features plus center frequency calculated across all frequency bands show distinctly separate means from the feature sets containing center frequency across the Alpha and Theta frequency bands. Specifically, the use of traditional spectral feature sets both alone, and in conjunction with center frequency across all bands demonstrated superior performance, producing the overall highest classification accuracy. Center frequency across Alpha and Theta bands had the poorest classifying ability and thus, is not considered an ideal feature set. The reason for this may be that limiting the input data to two frequency bands does not provide enough discriminating information for the SVM. Additionally, visual inspection of the PSD for each frequency band averaged across all participants and runs indicates no notable change in Alpha with the changing workload (Figure 13). As such, there will be minimal shifting of the center frequency due to changes in Alpha, thus the classifier is only able to extract meaningful information from the center frequency shifts in the Theta band, which can be noisy. Further, from the Tukey HSD analysis, for the EEG - derived feature sets containing traditional spectral features and spectral features plus center frequency calculated across all frequency bands, there is significant overlap of the means within the group, indicating a lack of distinct improvement of classification accuracy by adding center frequency across all bands to the traditional spectral feature set. This may be due to the presence of a ceiling effect that is masking the additive potential of center frequency.

Figure 13 Plot of power in the Alpha band calculated across all subjects and all runs for T5 Alpha shows a lack of notable difference in the underlying Alpha signal following a change in task difficulty.



Tukey HSD mean comparison analysis results indicate significant separation of means between the two transition types for the feature set containing only center frequency calculated across all frequency bands. Additionally, for the case where the workload level transitions from hard to easy, the mean classification accuracy using center frequency all bands shows significant overlap with the means for the spectral feature sets. As a result, center frequency across all bands is an equally useful parameter for classifying workload as the traditional spectral features when the workload level transitions from hard to easy. Unfortunately, the strength of center frequency as a single input metric appears to falter for the transition direction easy to hard workload, displaying significantly lower classification accuracies. The Tukey HSD confirms this observation as the mean classification accuracy using only center frequency all bands is distinctly lower for the case where workload transitions from easy to hard, with no significant overlap, than the mean classification accuracy for the case where workload transitions from hard to easy.

Supporting the observation of asymmetry between outputs based on transition direction, the ANOVA analysis indicates a significant interaction effect present in the data for the linear SVM but not the non – linear SVM. Generally, regardless of classifier type or feature set, there is a trend towards higher classification accuracies when the transition direction is from hard to easy as compared to easy to hard, although the difference is not always significant. More specific statistical analysis of the linear SVM output interaction effect reveals an asymmetry with respect to transition direction for the center frequency all bands feature set that is significant enough to drive the interaction effect. The linear SVM classifies workload using the center frequency all bands feature set with a significantly higher accuracy, with no overlap of means in the Tukey HSD analysis, when the workload changes from hard to easy as compared to the transition from easy to hard. The reason for this is not known. One explanation is, because the linear classifier attempts to create a line of maximal separation rather than a non – linear plane between the two classes, it may be more sensitive to hysteresis effects due to transition directions in the center frequency calculation. Additionally, the non – linear classifier may be less able to generalize to a new set of test data, resulting in over – fitting and reduced classification accuracies.

There appears to be physiological hysteresis in several frequency bands. Lower classification accuracy for the easy to hard condition fundamentally indicates a lack of distinct class differences. One factor affecting the class distinction is noise, another is slow physiological response. For the EEG frequency bands where the mean power response is clean and sharp, with minimal signal overlap following a transition from one class to another, as likely occurs in the transition from hard to easy, the classifier is more able to correctly assign workload levels to any given point, thus producing a higher accuracy output. Conversely, when the mean power response is slow and noisy following a transition from one class to another, as is likely the case for the easy to hard transition, the classifier is more likely to mistake one class for another, resulting in lower accuracies.

7. Conclusion

Results confirmed that spectral data and spectral plus center frequency are the most robust feature set, while center frequency across all bands is equally reliable for classifying workload in the case where cognitive workload level transitions from hard to easy. There is also evidence of physiological signal asymmetry based on transition direction. From the results, an average classification accuracy of 77 – 77.5 percent was achieved using the feature set containing traditional spectral features, an average classification accuracy of 77 – 79 percent was achieved with the spectral plus center frequency across all bands feature set, an average classification accuracy of 66 – 78 percent was achieved using center frequency calculated across all bands and an average classification accuracy of 63 – 68 percent was achieved using the center frequency calculated across the Alpha and Theta frequency bands.

Overall, the linear SVM classifier produced slightly higher classification accuracies than the non – linear SVM classifier, however the linear classifier was more sensitive to hysteresis. The results obtained from the classifier using the spectral feature sets were consistent with previous studies and the addition of center frequency calculated across all bands did not significantly improve classification accuracy. This result may indicate a degree of coinciding information between the spectral and center frequency all bands features rather than the addition of unique information. It may also indicate the presence of a ceiling effect with maximum classification accuracy. The center frequency calculated across Alpha and Theta bands alone as well as center frequency across all bands for the easy to hard condition displayed the poorest classification abilities among the eight feature set and transition types. It is possible the EEG features containing only center frequency do not contain enough information to classify workload or there are indiscriminate changes in the underlying signal which are not easily picked up by the classifier.

8. Directions for Future Work

This work presents several potential research directions for physiology – based operator based functional state assessment. Further analysis is needed to completely reveal the physiological cause of the transition type asymmetry. Investigation into the stability of center frequency as an independent feature may be advantageous, as limiting the input to a few features rather than the full traditional EEG feature set would reduce over – fitting. Additionally, determining features which are more uniquely correlated with EEG workload level, improving the strength of the classifier and delving deeper into potential physiological hysteresis effects that may be present would be appropriate continuation of this work. On the whole, this study presents novel concepts for neurological state assessment, helps to uncover underlying trends in EEG response to changing workload and furthers previous research for finding a practical method for classification of workload in modern aircraft systems.

References

Barnes, A. (1993). Instantaneous Spectral Bandwidth and Dominant Frequency with Aplications to Seismic Reflection Data. *Geophysics*, 419-428.

Berka, C., Levendowski, D., Lumicao, M., Yau, A., Davis, G., Zivkovic, V., et al. (2007). EEG Correlates of Task Engagement and Mental Workload in Vigilance, Learning and Memory Tasks. *Aviation, Space and Environmental Medicine*, B231 - B244.

Brookings, J., Wilson, G., & Swain, C. (1996). Psychophysiological responses to changes in workload during simulated air traffic control. *Biological Psychology*, 361 - 377.

Cannon, J., Krokhmal, P., Chen, Y., & Murphey, R. (2011). Detection of Temporal Changes in Physchophysiological Data Using Statistical Process Control Methods. *Computer Methods and Programs in Biomedicine*.

Christensen, J., Estepp, J., Wilson, G., & Russell, C. (2012). The effects of day - to - day variability of physiological data on operator functional state classification. *NeuroImage*, 57 - 63.

Comstock, J., & Arnegard, R. (1992). *The Multi-Attribute Task Battery for Human Operator Workload and Strategic Behavior Research.* Hampton: National Aeronautic and Space Administration.

Cox - Fuenzalida, L., Beeler, C., & Sohl, L. (2006). Workload History Effects: A Comparison of Sudden Increases and Decreases on Performance. *Social*, 8 - 14.

Crossen, S. (2011). Investigation of Variability in Cognitive State Assessment Based on Electroencephalogram - Derived Features. Dayton, OH: Wright State University.

Cumming, R., & Croft, P. (1973). Human Information Processing Under Varying Task Demand. *Ergonomics*, 581 - 586.

Fairclough, S., Venables, L., & Tattersall, A. (2005). The Influence of Task Demand and Learning on the Psychophysiological Response. *International Journal of Psychophysiology*, 171 - 184.

Farrell, P. (1999). The Hysteresis Effect. *Human Factors*, 226 - 240.

Fournier, L., Wilson, G., & Swain, C. (1999). Electrophysiological, behavioral and subjective idexes of workload when performing multiple tasks: manipulations of task difficuly and training. *International Journal of Psychophysiology*, 129 - 145.

Gevins, A., & Smith, M. (2005). Assessing Fitness - for - Duty and Predicting Performance with Cognative Neurophysiological Measures. *Biomonitoring for Physiological and Cognative Performance During Military Operations*.

Gevins, A., Smith, M., Leong, H., McEvoy, L., Whitfield, S., Du, R., et al. (1998). Monitoring Workload Menory Load during Computer - Based Tasks with EEG Pattern Recognition Methods. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 79 - 91.

Gevins, A., Smith, M., McEvoy, L., & Yu, D. (1997). High Resolution EEG Mapping of Cortical Activation Related to Working Memory: Effects of Task Difficulty, Type of Processing and Practice. *Cerebral Cortex*, 374 - 385.

Gevins, A., Zeitlin, G., Doyle, J., Yingling, C., Schaffer, R., Callaway, E., et al. (1979). Electroencephalogram Correlates of Higher Cortical Functions. *Science*, 665 - 668.

Gordon, J., & Grey, J. (1978). Perceptual effects of spectral modifications on musical timbres. *Journal of the Acoustical Society of America* , 1493 - 1500.

Grundel, A., & Wilson, G. (1992). Topographical Changes in the Ongoing EEG Related to the Difficulty of Mental Tasks. *Brain Topography*, 17 - 25.

Hankins, T., & Wilson, G. (1998). A Comparision of Heart Rate, Eye Activity, EEG and Subjective Measures of Pilot Mental Workload During Flight. *Aviation, Space and Environmental Medicine*, 360 - 367.

Helton, W., Shaw, T. W., J., M. G., Dember, W., & Hancock, P. (2004). Demand Transitions in Vigilance: Effects on Performance Efficency and Stress. *Human Performance, situation awareness and automation: Current research and trends*, 258 - 262.

Jung, T., Makeig, S., Stensmo, M., & Sejnowski, T. (1997). Estimating Alertness from the EEG Power Spectrum. *IEEE Transactions on Biomedical Engineering*, 60 - 69.

Kohlmorgen, J., Dornhege, G., Braun, M., Blankertz, B., Muller, K., Curio, G., et al. (2007). Improving Human Performance in a Real Operating Environment through Real - Time Mental Workload Detection. In *Toward Brain - Computer Interfacing* (pp. 409 - 422). Cambridge: MIT Press.

Krizo, E., Wilson, G., & Russell, C. (2005). NuWAM: The Development of a Modular, Real - Time Augmented Cognition System. *Proceedings of the 11th International Conference on Human - Computer Interaction.* Mahwah: Foundation of Augmented Cognition.

Krulewitz, J., Warm, J., & Wohl, T. (1975). Effects of shifts in the rate of repetitive stimulation on sustained attention. *Perception & Psychophysics*, 245 - 249.

Laine, T., Bauer, K., Lanning, J., Russell, C., & Wilson, G. (2002). Selection of Input Features Across Subjects for Classifying Crewmember Workload Using Artifical Neural Networks. *IEEE Transactions on Systems, Man and Cybernetics - Part A: Systems and Humans*, 691 - 704.

Lemm, S., Blankertz, B., Dickhaus, T., & Muller, K. (2011). Introduction to Machine Learning for Brain Imaging. *Neuroimage*, 387 - 399.

Ling, C., Goins, H., Ntuen, A., & Li, R. (2001). EEG Signal Analysis for Human Workload Classification. *IEEE*, 123 - 130.

Massar, M., Fickus, M., Bryan, E., Petkie, D., & Terzuoli, A. (2010). Fast Computation of Spectral Centroids. *Adv. Comput. Math*, 83 - 97.

Monnin, J., & Estepp, J. (2011). Contribution of High Frequency EEG Features to Physiologically - Based Operator Workload Estimation.

Moroney, B., Warm, J., & Dember, W. (1995). Effects of Demand Transitions on Vigilance Performance and Percieved Workload. *Proceedings of the Human Factors and Ergonomics Society 39th Annual Meeting*, (pp. 1375 -1379).

Ngoc Le, P., Ambikairajah, E., Epps, J., Sethu, V., & Choi, E. (2011). Investigation of spectral centroid features for cognative load classification. *Speech Communication*, 540 - 551.

Parasuraman, R., & Wilson, G. (2008). Putting the Brain to Work: Neuroergonomics Past, Present, and Future. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 468 - 474.

Peeters, G. (2004). *A large set of audio features for sound description (similarity and classification) in the CUIDADO project.* Paris, France.

Pereira, F., Mitchell, T., & Botvinick, M. (2009). Machine Learning Classifiers and fMRI: A Tutorial Overview. *Neuroimage*, S199 - S209.

Petrofsky, J., & Phillips, C. (1985). Discharge Characteristics of Motor Units and the Surface EMG During Fatiguing Isometric Contractions at Submaximal Tensions. *Aviation Space Environmental Medicine*, 581-586.

Phillips, C. (2000). *Human Factors Engineering.* New York: John Wiley and Sons.

Russell, C., & Gustafason, S. (2001). *Selecting Salient Features of Psychophysiological Measures.* Dayton: United States Air Force Research Lab.

Sulaiman, N., Taib, M., Aris, S., Hamid, N., Lias, S., & Murat, Z. (2010). Stress Features Identification from EEG Signals Using EEG Asymmerty and Spectral Centroids Techniques. *IEEE EMBS Conference on Biomedical Engineering and Sciences* (pp. 417 - 421). Kuala Lumpur: IEEE.

Sun, S., & Zhang, C. (2005). Assessing Features for Electroencephalographic Signal Categorization. *International Conference on Acoustics, Speech, and Signal Processing* (pp. V-417 - V-420). Philadelphia: IEEE.

Trejo, L., Knuth, K., Prado, R., Rosipal, R., Kubitz, K., Kochavi, R., et al. (2007). EEG - Based Estimation of Mental Fatigue: Convergent Evidence for a Three - State Model. In D. R. Schmorrow, *Augmented Cognition* (pp. 201 - 211). Heidelberg: Springer - Verlag.

Wilmore, J., & Costill, D. (2004). *Physiology of Sport and Exercise.* Champaign, IL: Human Kinetics.

Wilson, G. (1999). Operator Functional State Classification Using Neural Networks with Combined Physiological and Performance Features. *Proceedings of the Human Factors and Ergonomics Society 43rd Annual Meeting*, (pp. 1099 - 1102).

Wilson, G. (2002). An Analysis of Mental Workload in Pilots During Flight Using Multiple Psychophysiological Measures. *International Journal of Aviation Psychology*, 3 - 18.

Wilson, G., & Eggemeier, T. (1991). Psychophysiological assessment of workload in multi-task environments. In D. Damos, *Multiple Task Performance* (pp. 328-360). Taylor and Francis Ltd.

Wilson, G., & Fisher, F. (1995). Cognitive Task Classification Based Upon Topographic EEG Data. *Biological Psychology*, 239 - 250.

Wilson, G., & Russell, C. (2003). Real-Time Assessment of Mental Workload Using Psychophysiological Measures and Artificial Neural Networks. *Human Factors*, 635-643.

Wilson, G., Swain, C., & Ullsperger, P. (1999). EEG Power Changes During a Multiple Level Memory Retention Task. *International Journal of Psychophysiology*, 107 - 118.

Appendix A: Acronyms and Symbols

SVM – Support Vector Machine

EEG – Electroencephalogram

MATB – Multi – Attribute Task Battery

NASA – National Aeronautics and Space Administration

WPAFB – Wright – Patterson Air Force Base

EOG – Electrooculogram

EMG – Electromyogram

FFT – Fast Fourier Transform

CF – Center Frequency

HSD – Honestly Significant Difference

RM – Repeated Measures

RAS – Reticular Activating System

PSD – Power Spectral Density

NuWAM – New Workload Assessment Monitor

CA – Classification Accuracy

T – Transition Type (Easy to Hard or Hard to Easy)

F – Feature Set Type (Spectral, Spectral + CF, CF All Bands, CF Alpha/Theta Band)

n_{test} – Number of points in test set

$I(f(x_i),y)$ – Class Label Prediction Outcome (binomial)